LBL-12060
Preprint

Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

Submitted to Nuclear Instruments and Methods

THE MAXIMUM LIKELIHOOD METHOD APPLIED TO THE PROBLEM OF MULTI-COMPONENT DECAY

R.C. Eggers and L.P. Somerville

January 1981

NUCLEAR SCIENCE DIVISION

Prepared for the U.S. Department of Energy under Contract W-7405-ENG-48

The Maximum Likelihood Method Applied to the Problem of Multi-Component Decay

R. C. Eggers* and L. P. Somerville
Lawrence Berkeley Laboratory
University of California, Berkeley, CA 94720

Abstract

A method of multi-component decay analysis is presented using the maximum likelihood method instead of the least-squares method. The maximum likelihood method has been applied to the analysis of fission track data from experiments designed to produce elements 104-107, and has resulted in the best possible half-life estimates from data having only a few counts. It may also be applied to other cases where there are insufficient numbers of counts to make collection of data into time bins and least-squares analysis a valid procedure. The formulas for asymmetric error bars in the multi-component case are developed along with example computations.

Introduction

Multi-component decay problems are most often analyzed by the least-squares method. The least-squares method is derived from the Maximum Likelihood Method (MLM) under the assumption that there are a sufficient number of counts in each bin at all of the points where the data and the fitting function are compared. When this assumption breaks down, one must return to the MLM to obtain statistically correct results.

Some authors have published methods related to the MLM for the case of one component.[1,2] Recently, Zlokazov has given a rather complete description of this method.[3] He advocates reducing the problem to a single parameter. We find, in contrast, that there are many cases in which this cannot be done. Thus, a computer code is required which can handle an arbitrary number of components in the decay with the flexibility of fixing any of the degrees of freedom which may be known *a priori*. Zlokazov also defines a new error estimate somewhat related to the usual statistical variance which allows him to give asymmetric error bars. We will extend this into a formula for the case of many degrees of freedom.

*Present address: Cyclotron Corporation, Berkeley, CA 94710

Maximum Likelihood Problem Formulation

The likelihood function L is the product of the decay probability expressions for each time t_i that a count was detected. It is this function we wish to maximize with respect to all free parameters (half-lives and fractional number of atoms of each component at time zero). Our expression for the likelihood function is

$$\ln\left[L(N_j,\lambda_j)\right] = \sum_{i=1}^{n} \ln\left(\frac{\sum_{j=1}^{m} \lambda_j N_j e^{-\lambda_j t_i}}{\sum_{j=1}^{m} \sum_{k=1}^{\ell} \int_{t_{1k}}^{t_{2k}} \lambda_j N_j e^{-\lambda_j t}\, dt}\right), \tag{1}$$

where N_j is the fraction of the total counts in the decay of the jth component, and λ_j is the decay constant for the jth component. The experiment is assumed to have taken place such that it is sensitive to time intervals from t_{1k} to t_{2k}. (In practice this corresponds to segments of track detectors). Our expression differs from that of Zlokazov only in the normalizing denominator, which accounts for the fact that we may not observe all the counts associated with the decay. This denominator reduces the number of degrees of freedom of the likelihood function because if all the N_j are multiplied by the same constant, the same likelihood function value results. In order to achieve a system without degeneracy, one of the variables is eliminated from the equations by the use of the likelihood normalization formula

$$\sum_{j=1}^{m} \sum_{k=1}^{\ell} \int_{t_{1k}}^{t_{2k}} \lambda_j N_j e^{-\lambda_j t}\, dt = 1. \tag{2}$$

The eliminated N_j is called N_{j*}. The maximization of the likelihood function with this constraint is carried out with a multi-dimensional Newton's method. This then yields the most likely values for each of the parameters N_j and λ_j. The detailed formulas, including the first and second partial derivatives with respect to all the parameters, are available from the authors on request.

Once we have obtained the most likely values for the parameters, we want to calculate estimates for the errors. The definition of the variance for a parameter p_j (either an N_j or a λ_j shifted to a most likely value of zero) is

$$\sigma_{p_j^2} \equiv \frac{\int d\vec{p}\ p_j^2\ L(\vec{p})}{\int d\vec{p}\ L(\vec{p})} . \tag{3}$$

The evaluation of such integrals is quite time consuming and can be avoided by assuming that the likelihood function has a Gaussian shape in the region of the maximum. Under this assumption, we can generate estimates of the variance for all parameters from the second derivative matrix for the logarithm of the likelihood function. This can be seen from the following. The Gaussian shape of the maximum likelihood function can be expressed as

$$L(\vec{p}) = L_{max} \exp(-\tfrac{1}{2}\vec{p}^{\dagger}\underset{\sim}{\sigma}^{-1}\vec{p}), \tag{4}$$

where the matrix $\underset{\sim}{\sigma}^{-1}$ is an arbitrary designation for the second derivative matrix which gives the likelihood function the correct shape, and $\vec{p}^{\dagger}$ is the transpose of the column vector $\vec{p}$. We intend to show that its inverse $\underset{\sim}{\sigma}$ is actually the variance-covariance matrix as defined by

$$\sigma_{p_i p_j} \equiv \sigma_{ij} \equiv \int d\vec{p}\ p_i p_j\ L(\vec{p}). \tag{5}$$

Now consider the multi-dimensional integral

$$G = \iint \ldots \int dp_1 \ldots dp_{i-1} dp_{i+1} \ldots dp_m dp_i p_i \sum_{j=1}^{m} (\underset{\sim}{\sigma}^{-1})_{ij} p_j \exp(-\tfrac{1}{2}\vec{p}^{\dagger}\underset{\sim}{\sigma}^{-1}\vec{p}). \tag{6}$$

The integration is carried out over p_i. Integrating by parts, we differentiate the p_i and integrate the exponential factor with the remainder of the pre-exponential factor. This gives

$$G = \int d\vec{p}\ \exp(-\tfrac{1}{2}\vec{p}^{\dagger}\underset{\sim}{\sigma}^{-1}\ \vec{p}). \tag{7}$$

Next, consider a second type of multi-dimensional integral

$$H = \iint \ldots \int dp_1 \ldots dp_{i-1} dp_{i+1} \ldots dp_m dp_i p_k \sum_{j=1}^{m} (\underset{\sim}{\sigma}^{-1})_{ij} p_j \exp(-\tfrac{1}{2}\vec{p}^{\dagger}\underset{\sim}{\sigma}^{-1}\vec{p}) \tag{8}$$

with $i \neq k$. We can integrate this directly over dp_i, since p_k (with $i \neq k$) is a constant with respect to integration over p_i. We then have

$$H = \int\int\ldots\int dp_1\ldots dp_{i-1}dp_{i+1}\ldots dp_m p_k \exp(-\tfrac{1}{2}\vec{p}^\dagger \underset{\sim}{\sigma}^{-1}\vec{p})\Bigg|_{p_i=-\infty}^{p_i=+\infty} = 0. \qquad (9)$$

These two types of integrals are sufficient to prove the following matrix equation correct, since each element of the matrix on the right hand side corresponds to one of the two types of integrals.

$$\begin{pmatrix} \sigma_{11}^{-1} & \sigma_{12}^{-1}\cdots\sigma_{1m}^{-1} \\ \sigma_{21}^{-1} & \sigma_{22}^{-1}\;\cdots\cdots \\ \vdots & \\ \sigma_{m1}^{-1}\;\cdots & \sigma_{mm}^{-1} \end{pmatrix} \begin{pmatrix} \int p_1p_1e^{-\frac{1}{2}\vec{p}^\dagger \underset{\sim}{\sigma}^{-1}\vec{p}}d\vec{p} & \int p_1p_2e^{-\frac{1}{2}\vec{p}^\dagger \underset{\sim}{\sigma}^{-1}\vec{p}}d\vec{p}\ldots & \int p_1p_me^{-\frac{1}{2}\vec{p}^\dagger \underset{\sim}{\sigma}^{-1}\vec{p}}d\vec{p} \\ \int p_2p_1e^{-\frac{1}{2}\vec{p}^\dagger \underset{\sim}{\sigma}^{-1}\vec{p}}d\vec{p} & \int p_2p_2e^{-\frac{1}{2}\vec{p}^\dagger \underset{\sim}{\sigma}^{-1}\vec{p}}d\vec{p} & \cdots\cdots\cdots \\ \vdots & & \\ \int p_mp_1e^{-\frac{1}{2}\vec{p}^\dagger \underset{\sim}{\sigma}^{-1}\vec{p}}d\vec{p} & \cdots\cdots\cdots & \int p_mp_me^{-\frac{1}{2}\vec{p}^\dagger \underset{\sim}{\sigma}^{-1}\vec{p}}d\vec{p} \end{pmatrix}$$

$$\times \frac{1}{\int e^{-\frac{1}{2}\vec{p}^\dagger \underset{\sim}{\sigma}^{-1}\vec{p}}d\vec{p}} = I, \qquad (10)$$

where I is the identity matrix.

Thus, these two matrices are inverses of each other as required. In our code LIKELY, which is based on the MLM, we generate error bars assuming the likelihood function to be Gaussian-like in the region of the maximum.

Figure 1 is an application of LIKELY to a six parameter analysis of some spontaneous fission data in search of element 104 isotopes. To check the assumption that the maximum is Gaussian-like, figure 2 is a plot of the likelihood function in the region of the maximum for six parameters. The profile for N_{j*} (which has been eliminated from the equations) was generated based on the assumed direction in parameter space for $\hat{N}_{j*}$.

$$\hat{N}_{j*} = \frac{1}{\alpha}\sum_j \frac{\partial N_{j*}}{\partial \lambda_j}\hat{\lambda}_j + \frac{1}{\alpha}\sum_{\substack{j \\ j\neq j*}} \frac{\partial N_{j*}}{\partial N_j}\hat{N}_j\ , \qquad (11)$$

where

$$\alpha = \left(\frac{\partial N_{j*}}{\partial \lambda_j}\right)^2 + \sum_{\substack{j \\ j\neq j*}} \left(\frac{\partial N_{j*}}{\partial N_j}\right)^2 . \qquad (12)$$

The program also calculates a goodness of fit parameter, χ^2, based on an adaptive binning procedure in which the time axis is divided into bins, each of which has roughly an equal number of counts.

Formulation of Asymmetric Error Bar Problem

For the one-dimensional case, Zlokazov,[3] in his eq. (9), defined a relationship between the upper and lower half-life limits $t_{\frac{1}{2},u}$ and $t_{\frac{1}{2},\ell}$ and the confidence level f as follows:

$$\frac{\int_{t_{\frac{1}{2},\ell}}^{t_{\frac{1}{2},u}} L(t_{\frac{1}{2}})\, dt_{\frac{1}{2}}}{\int_0^\infty L(t_{\frac{1}{2}})\, dt_{\frac{1}{2}}} = f_{t_{\frac{1}{2}}} \tag{13}$$

with $L(t_{\frac{1}{2},u}) = L(t_{\frac{1}{2},\ell})$. (14)

We choose f to be 0.9. The resulting confidence limits $t_{\frac{1}{2},\ell}$ and $t_{\frac{1}{2},u}$ are related to the width of the likelihood distribution and may be asymmetric as measured from the point of maximum likelihood.

If there is more than one free parameter, we have to apply the more general formula

$$\frac{\int_{p_{i\ell}}^{p_{iu}} [\int\int \ldots \int L(\vec{p})\, dp_1 \ldots dp_{i-1} dp_{i+1} \ldots dp_m]\, dp_i}{\int d\vec{p}\, L(\vec{p})} = f_{p_i} \tag{15}$$

with $L(p_{iu}) = L(p_{i\ell})$. (16)

We have to perform the p_i integration on the projection of the likelihood function on the p_i axis (expression in brackets, equation 15) instead of on the likelihood function itself. Otherwise, we would ignore off-axis likelihood contributions. The integral of the likelihood function along the p_i axis gives a severely underestimated width for the likelihood. The variance of $\sigma_{p_i^2}$ does include the off-axis contributions because it is defined as

$$\sigma_{p_i^2} \equiv \frac{\int d\vec{p}\, p_i^2\, L(\vec{p})}{\int d\vec{p}\, L(\vec{p})} . \tag{17}$$

Adopting the approach of equations (15) and (16), we see that we have to solve a problem in m-dimensional numerical integration, where m is the number of free parameters in the problem. The reason for the numerical approach is that the likelihood function projections will not yield to analytical integration, since the Gaussian part of the expressions will not. We have therefore written a code called BUMP which evaluates formulas (15) and (16) after LIKELY locates the maximum likelihood point and passes on the parameters to BUMP so BUMP can delimit the region of interest for the integration. Since we are dealing with repeated numerical integrations, the process is quite slow. A case of three free parameters took approximately 1 hour of computation time on a PDP-11/60. Thus, we have to choose examples carefully.

Figure 3 is an application of LIKELY to some other data from element 104 experiments with three parameters. Figure 4, which was derived from BUMP, shows the projection of the likelihood functions on the axis of each of the three free parameters. Note in table 1 that approximate σ values calculated from LIKELY, along with upper and lower limits calculated from BUMP, agree within a few percent.

Summary and Conclusions

When working with mica track detectors, where the only distinctions between various products are different half-lives, the importance of correct data analysis cannot be understated. The χ^2 value calculated by LIKELY should be of great help in deciding if the correct number of components has been chosen.

LIKELY is slower than least-squares programs by the factor of the number of counts in the bins for the least-squares programs. Still, it runs quite fast; for the data in figure 3, with ~1200 tracks, the program takes less than a minute to execute on a PDP-11/60. LIKELY results are 10 to 15 percent different from the least-squares programs (in cases like figure 1 and figure 3). On balance, we think that the MLM is the most appropriate procedure to treat track detector data.

The BUMP program, however, takes much computer time and, for the data in figure 3, shows only slight changes in the error bars, which indicates that it is not worthwhile running for every data set. Asymmetric errors may be more important as the number of data points decreases, but the time of calculation does not decrease significantly with fewer data points; it goes up about an order of magnitude for each additional degree of freedom.

Acknowledgements

The authors would like to thank A. Ghiorso and J. M. Nitschke for providing the data from element 104 experiments. We also wish to thank M. E. Leino for his assistance, and the Cyclotron Corporation, Berkeley, for many hours of computer time. This work was supported in part by the U. S. Department of Energy under Contract W-7405-ENG-48.

References

1. R. D. Evans, "Mean-Life Determination by Peierl's Method", The Atomic Nucleus, McGraw-Hill (1955) p. 812.
2. R. A. Fisher, Messenger Math. 41 (1912) p. 155.
3. V. B. Zlokazov, Nucl. Inst. and Meth. 151 (1978) pp. 303-306.
4. J. M. Nitschke et al., to be published in Nucl. Phys. (1981).

Figure Captions

Figure 1:

Maximum likelihood fit to the spontaneous fission data for the reaction 95 MeV ^{18}O + ^{248}Cm. For the purpose of illustrating the maximum likelihood code, the "3.1 ms spontaneous fission activity" has been included, although other experiments indicated that it is probably due to scattered beam.

Figure 2:

Profile of the likelihood function near the maximum for each variable (counts and half-life) for the data in figure 1. The components are: (1) 992 ms; (2) 3.14 ms; and (3) 53.7 ms.

Figure 3:

Maximum likelihood fit to the spontaneous fission decay curve for the reaction 81.6 MeV ^{15}N + ^{249}Bk.[4]

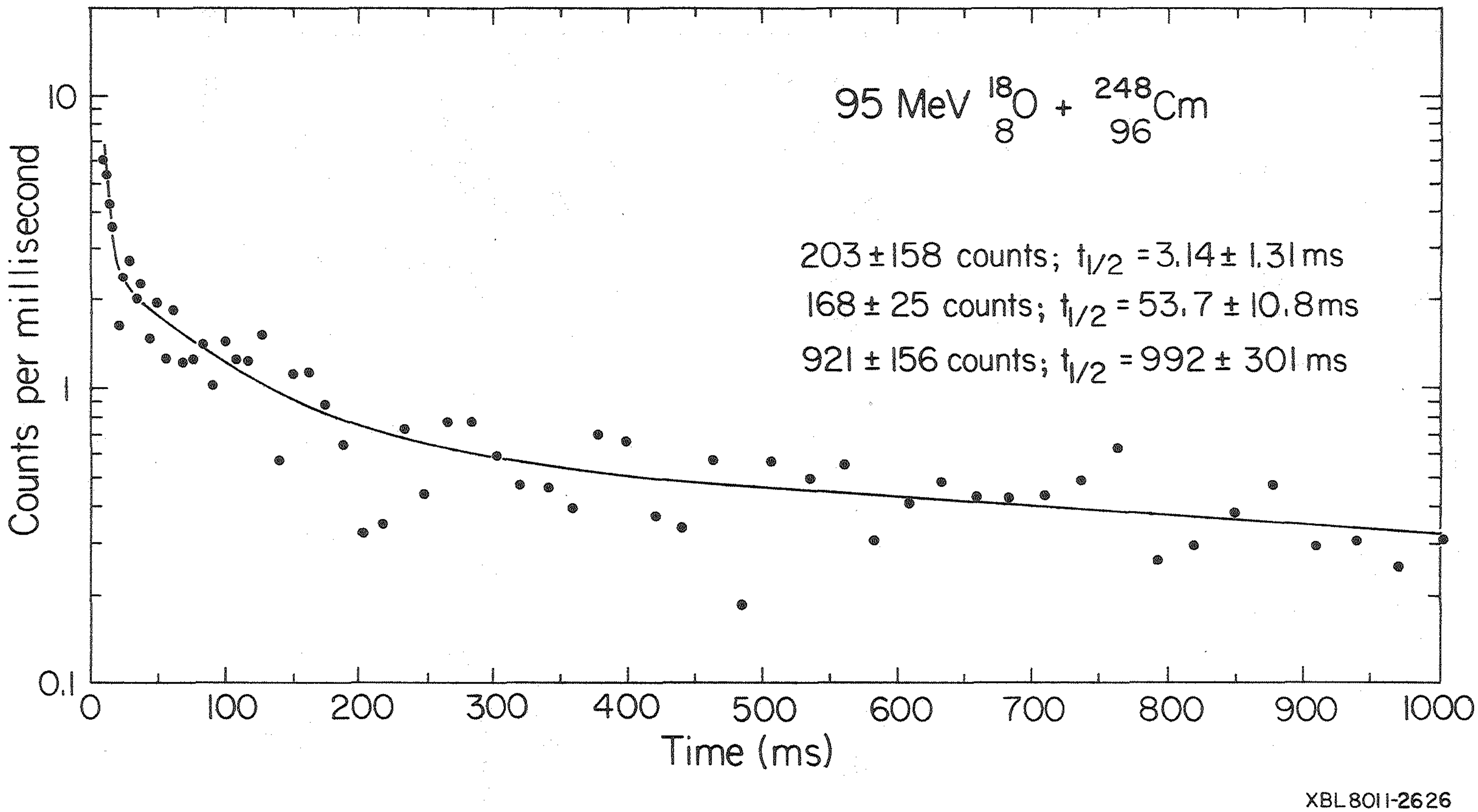
Figure 4:

Projections of the likelihood function on the parameter axes for the data in figure 3. The projections are seen to be roughly Gaussian.

Table 1

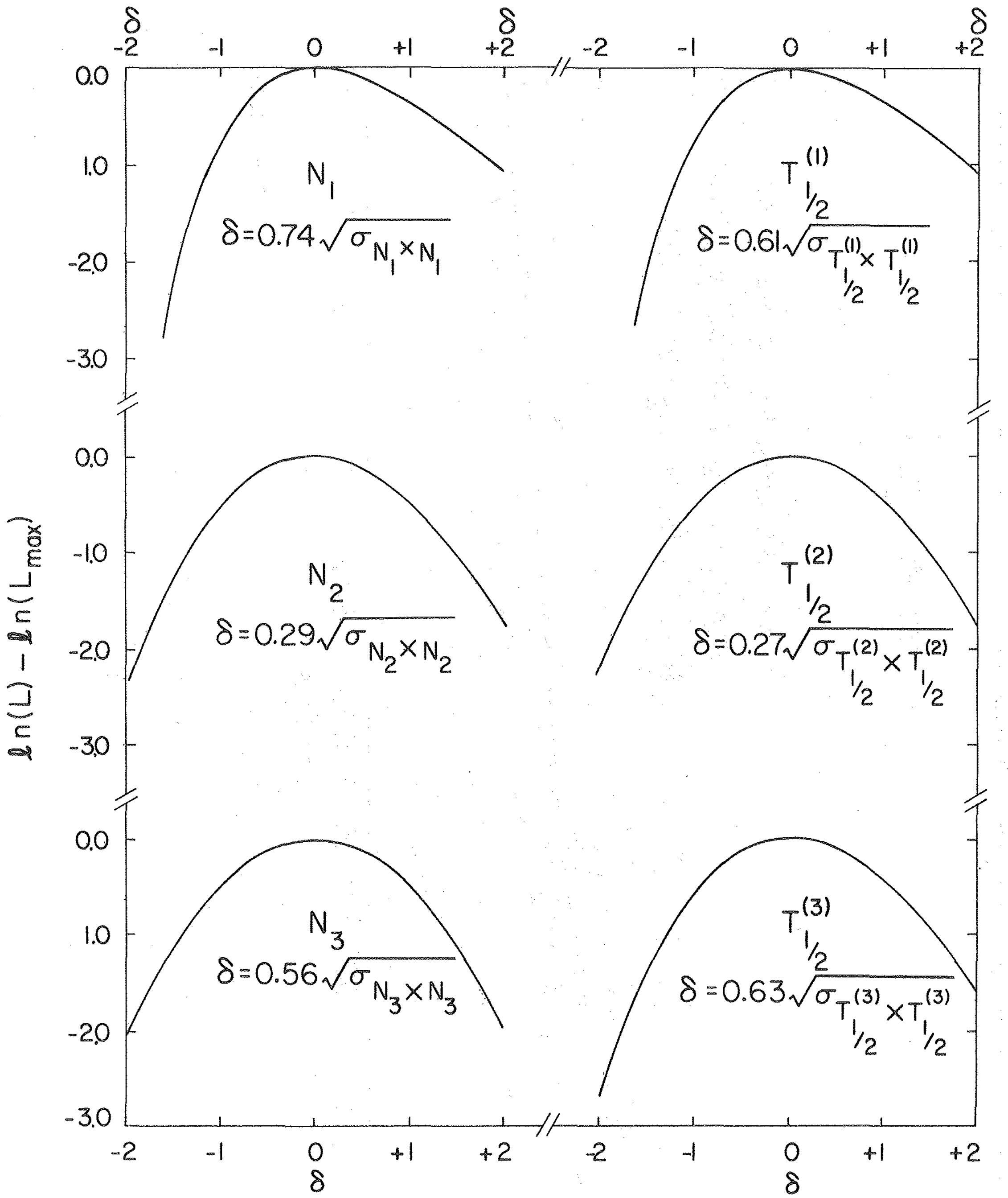
Comparison of "LIKELY" and "BUMP" Results for the data of figure 3

LIKELY	BUMP
$T_{\frac{1}{2}}^{(1)} = 23.29 \pm 1.27$ ms	$23.29 \begin{smallmatrix} +\ 1.32 \\ -\ 1.23 \end{smallmatrix}$ ms
$N_1 = 1{,}233 \pm 61.5$ counts	$1{,}233 \begin{smallmatrix} +\ 62.6 \\ -\ 60.3 \end{smallmatrix}$ counts
Background = 718.5 ± 29.5 counts	$718.5 \begin{smallmatrix} +\ 29.1 \\ -\ 30.0 \end{smallmatrix}$ counts



95 MeV $^{18}_{8}O + ^{248}_{96}Cm$
203 ± 158 counts; $t_{1/2}$ = 3.14 ± 1.31 ms
168 ± 25 counts; $t_{1/2}$ = 53.7 ± 10.8 ms
921 ± 156 counts; $t_{1/2}$ = 992 ± 301 ms
Counts per millisecond
Time (ms)
10
1
0.1
0
100
200
300
400
500
600
700
800
900
1000
XBL 8011-2626

Fig. 1



XBL 8012-2499

Fig. 2

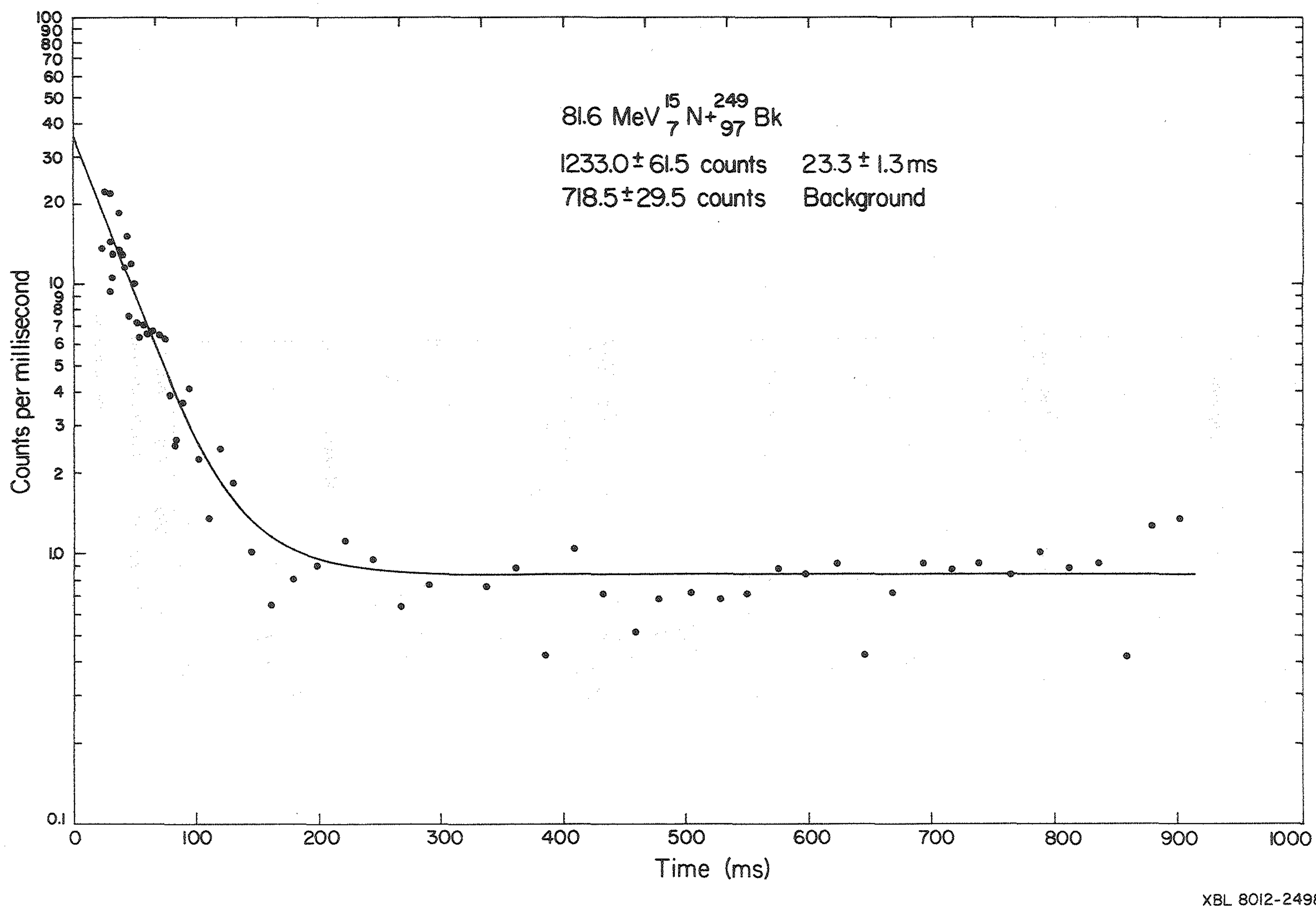

81.6 MeV $^{15}_{7}N + ^{249}_{97}Bk$
1233.0 ± 61.5 counts 23.3 ± 1.3 ms
718.5 ± 29.5 counts Background
Counts per millisecond
Time (ms)
XBL 8012-2498

Fig. 3

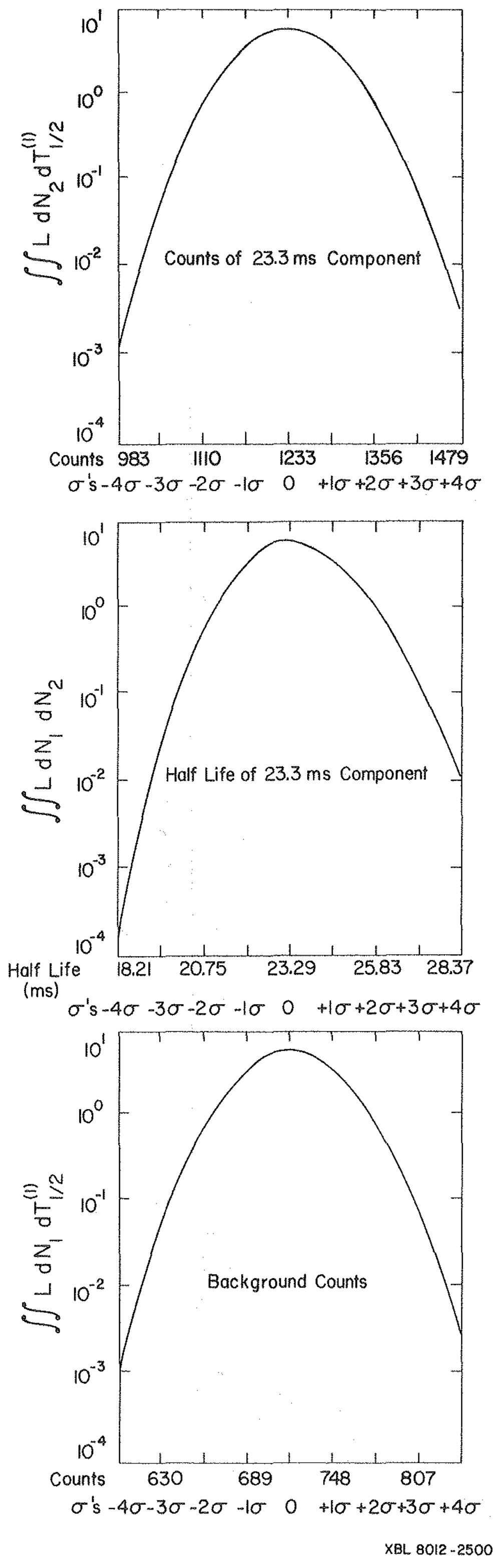

Counts of 23.3 ms Component
Counts 983 1110 1233 1356 1479
σ's -4σ -3σ -2σ -1σ 0 +1σ +2σ +3σ +4σ
Half Life of 23.3 ms Component
Half Life (ms) 18.21 20.75 23.29 25.83 28.37
σ's -4σ -3σ -2σ -1σ 0 +1σ +2σ +3σ +4σ
Background Counts
Counts 630 689 748 807
σ's -4σ -3σ -2σ -1σ 0 +1σ +2σ +3σ +4σ
XBL 8012-2500

Fig. 4